Cybersecurity

Login Analysis

Overview

Businesses that use the digital services of a Software-as-a Service (SaaS) company often require the SaaS company to provide assurance regarding the internal controls of the SaaS solution it is providing. Indeed, a business that uses a SaaS company for material processes to the financial statements is required by external auditing standards to test the internal controls of the SaaS company. Since SaaS companies can have many customers, rather than have each customer's external auditor examine the SaaS company, the SaaS company can obtain a System and Organization Controls (SOC) report.

As part of a SOC report, a SaaS company has an external auditor evaluate and opine on the internal controls of the SaaS solution leveraged by its business customers. The SOC report then allows all external auditors to rely on the procedures performed and documented in the report so they do not have to re-evaluate the SaaS company's controls. Of particular importance is a SOC Type 2 report wherein a CPA opines about internal controls related to security, availability, processing integrity, confidentiality or privacy. For more information about SOC reports, see https://us.aicpa.org/interestareas/frc/assuranceadvisoryservices/serviceorganization-smanagement).

As part of a SOC engagement, the external auditor will often perform internal control testing and analysis of various computer logs. A well-designed computer system will log interactions that users have with the system. For example, a computer system can be designed to create an entry in a log file anytime a user logs into the system, accesses certain information, downloads a file or has any other interaction with the system. The analysis of these log files is not only useful for SOC engagements, but it is often a critical part of cybersecurity audits, forensic analyses, security analyses and criminal investigations.

For this case, you will be performing tests related to authentication logs. An authentication log is a list of all attempts made to log in (i.e., authenticate) to a computer system. The data for this case is simulated. The problems listed below are intended to help you practice dealing with data and are patterned after tests you would perform as part of a cybersecurity engagement.

Please assume the company has established certain policies and procedures about logging into systems. The company requires users to authenticate to access the company's network and whenever they leave their computer (e.g., to use the restroom or go to a meeting), they must secure their machine and then log in again upon returning to their workstation. All company computers are set up so that if there is no interaction with the machine in a defined period of time, the computer automatically locks and requires the user to reauthenticate. Given these policies, it is common to have users log in many times during the day.

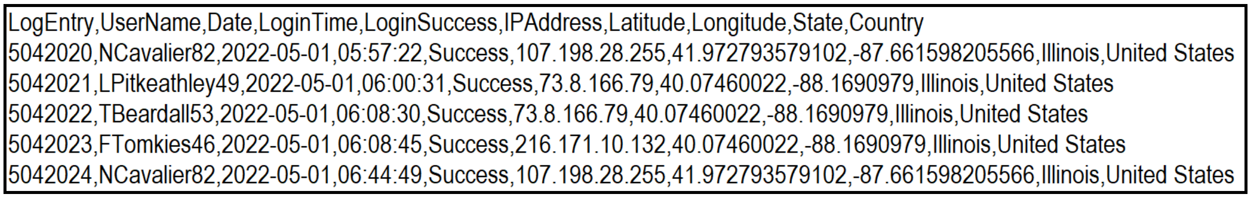
Data description

The following describes each of the data items included.

In the file **Cybersecurity_case_studies_LoginAnalysis_LogFile.csv**, the fields are defined as follows:

- LogEntry: A unique identifier for every entry. The system automatically assigns the next integer value to the next attempted login.
- UserName: The unique username generated for each employee.
- Date: The date of the attempted login in the format of YYYY-MM-DD.
- LoginTime: The time stamp of the attempted login using military time (i.e., 24 hours) in the format of HH:MM:SS.
- LoginSuccess: Identifies where the login attempt was a “success” or a “failure.”
- IPAddress: The IP address using IPv4 format of the attempted login.
- Latitude: The latitude associated with the IP address.
- Longitude: The longitude associated with the IP address.
- State: The US state associated with the latitude and longitude.
- Country: The country associated with the latitude and longitude. On the tab “PasswordFile,” the username and hashed password for each current employee are listed.

The file is comma delimited. A screenshot of a few rows of the data is presented below. Note that the file contains more records than can be displayed in Microsoft Excel. To view all records, consider using a program such as Notepad.

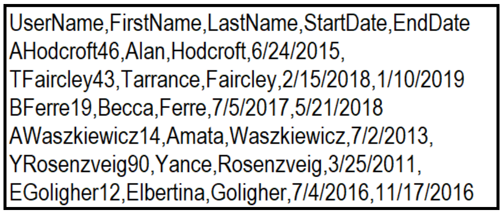


```
LogEntry,UserName,Date,LoginTime,LoginSuccess,IPAddress,Latitude,Longitude,State,Country
5042020,NCavalier82,2022-05-01,05:57:22,Success,107.198.28.255,41.972793579102,-87.661598205566,Illinois,United States
5042021,LPitkeathley49,2022-05-01,06:00:31,Success,73.8.166.79,40.07460022,-88.1690979,Illinois,United States
5042022,TBeardall53,2022-05-01,06:08:30,Success,73.8.166.79,40.07460022,-88.1690979,Illinois,United States
5042023,FTomkies46,2022-05-01,06:08:45,Success,216.171.10.132,40.07460022,-88.1690979,Illinois,United States
5042024,NCavalier82,2022-05-01,06:44:49,Success,107.198.28.255,41.972793579102,-87.661598205566,Illinois,United States
```

In the file **Cybersecurity_case_studies_LoginAnalysis_EmployeeData.csv**, the fields are defined as follows:

- UserName: The unique username generated for each employee.
- FirstName: The first name of each employee.
- LastName: The last name of each employee.
- StartDate: The date the employee was hired to begin working.
- EndDate: The last day the employee worked at the company.

The file is comma delimited. A screenshot of a few rows of the data is presented below.



```
UserName,FirstName,LastName,StartDate,EndDate
AHodcroft46,Alan,Hodcroft,6/24/2015,
TFaircley43,Tarrance,Faircley,2/15/2018,1/10/2019
BFerre19,Becca,Ferre,7/5/2017,5/21/2018
AWaszkiewicz14,Amata,Waszkiewicz,7/2/2013,
YRosenzveig90,Yance,Rosenzveig,3/25/2011,
EGoligher12,Elbertina,Goligher,7/4/2016,11/17/2016
```

Cybersecurity

Login Analysis

Required

- ► Complete the following requirements using Alteryx or the software system specified by your instructor.
- ► Submit your completed workflows in a packaged Alteryx workbook (.yxzp file type [Options > Export Workflow >]) saved with a naming convention to include your full name, e.g., Cybersecurity_case_studies_LoginAnalysis_FirstName_LastName.yxzp.
- ► Annotate each step in your workflow to indicate the purpose of that step. Use comments to indicate the part of the workflow that answers each question posed.

The case questions are separated into various sections. The first three questions have definitive right or wrong answers. The subsequent questions require you to exercise professional judgment to correctly answer them.

Note: For purposes of this case, assume the date is December 1, 2022.

Cybersecurity

Login Analysis

Part 1: Basic questions to become better acquainted with the data.

a. How many total login attempts were made in the data set?

b. What percentage of all logins failed (present answer as XX.XX%, round to two decimal places)?

c. What is the total number of login attempts made on each day of the week (i.e., Monday, Tuesday)? If using Alteryx, the DateTime functions can be helpful for this problem. Does the distribution of logins by day make sense? Why or why not?

d. What is the average number of login attempts an employee makes each day (round to two decimal places)?

Cybersecurity

Login Analysis

Part 2: Employee analysis

*HINT: Make sure you think carefully about how you will join the two data sets before answering these questions. You want to analyze **all** employees and **all** logins, even if there are null values in some fields*. *Remember null values are not treated as zeros. So, if you perform a calculation on a null value, the result will be a null value.*

a. List every employee (one row per employee) who has worked at the company and show the total number of successful logins and the total number of failed logins they have made. Your answer should have four columns labeled in the following order: UserName, EmployeeName (concatenated in one field: FirstName LastName), SuccessfulLogins and FailedLogins. Sort the data so the employees with the most successful logins are listed first.

b. Compute the failure login rate for each employee (computed as failed logins divided by the total login attempts) and sort the data to show employees who most frequently fail at logging in at the top of the list. Only show employees who have attempted at least one login in the data. Your final answer should have these columns in this order: UserName, EmployeeName (concatenated in one field: FirstName LastName) and FailureRate. What problems do you see from doing this analysis?

c. Identify employees who attempted to log in a day or longer (i.e., not the same day) after they were terminated from working at the organization. The final table should have the following columns in this order: LogEntry, EmployeeName, EndDate, LoginDate and LoginSuccess. Sort the data in ascending order by LogEntry. Based on this analysis, what recommendations would you make to the company to strengthen its internal controls?

d. Identify individuals listed in the employee file who have at least one login in the data, have not been terminated and have not logged in for 30 days or longer. Remember to assume that the date you are performing this case is December 1, 2022. The final table should have the following columns in this order: UserName, LastLoginDate and TimeToLogin. Sort the data in descending order by TimeToLogin. LastLoginDate is the date of the employee's last login attempt. The TimeToLogin variable should measure the number of days since the employee's last login attempt (this variable should be a positive value).

e. How many attempted logins were made that do not correspond to an employee listed in the employee file? How many of these attempts were successful vs. unsuccessful? Provide any insight into what you believe is happening with these logins.

Part 3: Data integrity

a. One control with log data is to include a field to uniquely identify every transaction. In the log file, the field LogEntry is used to keep track of every attempted login. The field auto increments for each new attempted login. As such, it should increase by one for every login attempt. Evaluate the data to determine if any log entry data is missing. How many records, if any, are missing?

b. Provide the LogEntry numbers for all of those that are missing. Sort the missing LogEntry numbers in ascending order.

c. Which individual or individuals do you believe was most likely to have deleted the login records? Justify your answer.

Part 4: Holidays

a. The company is closed on several US holidays. The holidays the company is closed included in the data set include Memorial Day (May 30, 2022), Independence Day (July 4, 2022), Labor Day (September 5, 2022) and Thanksgiving Day (November 24, 2022). Some employees may still work on these days, but it is expected that far fewer employees will log in.

   On these holidays, did the company experience fewer logins than what you would expect? Cleary state your expectation, how many logins occurred on each holiday and anything else of importance that you notice about logins on these four holidays.

Cybersecurity

Login Analysis

Part 5: Location

a. While employees can travel, there is a reasonable amount of travel possible within a certain time limit. Identify any logins that do not seem reasonable based on the distance the employee traveled relative to their previous login. You should not include any login attempts outside the United States for this question, and know that the latitude and longitude are approximate, so small differences may be explained by deficiencies in the data. Cleary state what your expectation is, why logins are not deemed reasonable and anything else of importance that you notice about the physical distance between the locations.