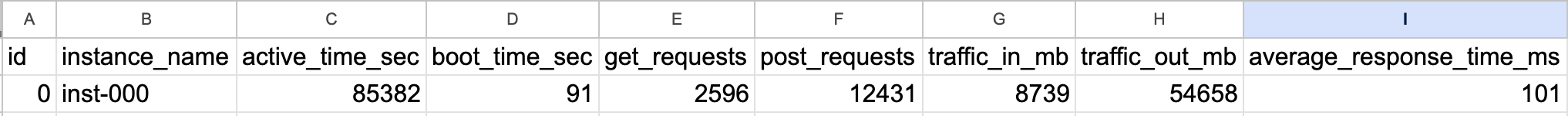
**CodeSignal Questions**

Data Analyst

**Scenario Questions**

1. Your platform develops a service that runs on a cluster of N servers in a single datacenter. Every request goes through load balancer that sends this request to one of N instances for processing. Every server in the cluster permanently collects and stores various internal information for the last 24 hours. Given this information as a dataset, answer the following questions.

Data Explanation: In the header and first row of data:=



During the last 24 hours, server ***inst-000*** was actively in work for ***85382*** seconds. It has also rebooted one or more time, which took ***91*** seconds in total. Note that booting time is not counted as active time work. The server has processed ***2596*** GET requests and ***12431***POST requests. According to traffic measures, the incoming traffic is ***8739*** Mb and the outgoing traffic is ***54658*** Mb. The average instance response time is ***101*** msec.

Data File to use: data1.csv

* 1. Let’s mark the server to be ***high-efficient*** if it has an average response time under ***150ms***, processed at least ***22000*** total requests, and has a boot time percentage under ***1%*** of active time. Count the number of ***high-efficient*** servers.
     1. Choices:
        1. 53
        2. 241
        3. 61
        4. **51 (Verified)**
     2. Solution Formula: = IF(AND(I2 < 150, SUM(E2, F2) >= 22000, D2 / C2 < 0.01), 1, 0)
  2. Let’s mark the server to be *“****uploader****”* if the incoming traffic is higher than outgoing traffic and the server has processed at least **1000** POST requests. Count the number of such servers.
     1. Choices
        1. 39
        2. 10 (Verified)
        3. 103
        4. 127
     2. Solution Formula: =IF(AND(G2 > H2, F2 >= 1000), 1, 0)
  3. Let’s assume that an inefficient server is a server with boot time above 0.3% of active working time, has less than 5 total active work hours, and has a response time greater than 150ms. What percentage of servers are considered to be inefficient?
     1. Choices
        1. 16.67%
        2. 38.00%
        3. 32.33%
        4. 1.67% (Verified)
     2. Solution Formula: IF(AND((D2 / C2) > 0.003, C2 < 18000, I2 > 150), 1, 0), take the sum of all the values and divide by total number of servers (300 in this example)
  4. How many servers are marked as *“****uploader****”* and NOT *“****high efficient****”*?
     1. Choices
        1. 95
        2. 57
        3. 6 (Verified)
        4. 47
     2. Solution Formula: Using result from a and b (assume columns K and L): IF(AND(K2 = 0, L2 = 1), 1, 0)
  5. Percent of boot time can be calculated as the boot time of a server over its total active work time. For example, for a server with a boot time of 5 sec and active work time 500 sec, boot time takes 1% of active work time of the instance. What is the average percent of boot time over the servers?
     1. Choices
        1. 0.98%
        2. 57.63%
        3. 0.48% (Verified)
        4. 48.12%
     2. Solution Formula: SUM((D2 / C2)) – sum all of these down the column and divide by 300 and take the percentage of result

**Coding Questions**

1. You own a small online store and want to analyze customer ratings for the products that you're selling. After doing a data pull, you have a list of products and a log of purchases. Within the purchase log, each record includes the number of stars (from 1 to 5) as a customer rating for your product.

Specifically you now have a database containing 2 tables - purchases and products. The purchases table contains the following columns:

* + id - a unique id for each purchase
  + product\_id - the id of the product purchased
  + stars - the number of stars (from 1 to 5) that the customer provided as a rating for the product

The products table contains:

* + id - unique id for the product
  + name - the name of the product
  + category - the category the product belongs to
  + price - the unit price of the specific product

Ror each category, find the lowest price among all products that received at least one 4-star rating from customers 2. If a product category did not have any products that received at least one 4-star rating, the lowest price is considered to be 0. 3. the final output should be sorted by product category in alphabetical order.

Ex.

Purchases

|  |  |  |
| --- | --- | --- |
| Id | Product\_id | stars |
| 1 | 1 | 2 |
| 2 | 3 | 3 |
| 3 | 2 | 2 |
| 4 | 4 | 4 |
| 5 | 6 | 5 |
| 6 | 6 | 4 |

Products

|  |  |  |  |
| --- | --- | --- | --- |
| Id | Name | Category | Price |
| 1 | Cripps Pink | Apple | 10 |
| 2 | Navel Orange | Orange | 12 |
| 3 | Golden Delicious | Apple | 6 |
| 4 | Clementine | Orange | 14 |
| 5 | Pinot Noir | Grape | 20 |
| 6 | Bing Cherries | orange | 36 |

The output should be

|  |  |
| --- | --- |
| Product\_category | Price |
| Apple | 0 |
| Cherry | 36 |
| Grape | 0 |
| orange | 14 |

Answer: Question1