CDS QV

Data Analyst Take-Home Test

Introduction

Hello and welcome to the first step in the hiring process! We sincerely thank you for your time and interest in this position and look forward to reviewing your work.

Here at EA's CDS Analytics (CDSA) team we work hard to support and improve Quality Verification (QV) practices across all of EA's game teams. Likewise, today you will be working with small sample of data similar to what our Data Analysts work with to answer a few foundational data analysis questions.

The Central Development Services (CDS) team at EA works directly with all game teams in support of providing fast and actionable feedback on quality during the development cycle of our video games. We are a distributed, cross-functional team, full of highly creative people who work in an environment that utilizes cutting-edge technologies in support of the game development process.

We hope that this exercise provides you with a better understanding of the type of work Data Analysts do within our team while also allowing us to better understand your methodology, thought processes, and overall fit for the position.

Suggested Timeline

This test has been designed to be completed in roughly 3 hours. As we are most interested in gaining an accurate understanding of your ability without the strict pressure of time, we provide all candidates with a full 48 hours to complete and submit their solutions.

We want to be careful to not demand too much of your time - please do not spend all 48 hours on this test. The test is intended to evaluate candidates for basic data analysis skills and, we promise, was not designed to include any "trick" questions.

Guidelines

Candidates are strictly prohibited from receiving outside help in the completion of this test. Any candidates found to have plagiarized or received help from a third party will be removed from the hiring process.

Candidates are strictly prohibited from distributing any materials provided in this process at any point in time. This includes these instructions, the CSV files containing data, and any analysis completed using the data.

Evaluation

We will be evaluating all candidates based on the below criteria:

- **Correctness**: Did you get the correct final answers?

- **Logic:** Is the logic in your answer sound?

- **Assumptions:** Did you make any assumptions clear?

- **Code Quality:** Is the code executable, tested, functional, documented?

- **Efficiency:** Is the code concise and reasonably performant?

- **Technology Used:** Are you using modern tools and libraries appropriately?

- **Communication:** Were answers clear and presented in a reasonable way?

Submission Details

Please submit your solutions using *a single PDF file* that *does not contain any personal identifiable information* (e.g. name, email, phone number, etc.) via email to dclarke@ea.com before **12:00 pm on Wednesday, June 12.**

We understand that applicants are used to working with a variety of analysis, visualization, and presentation programs and accept solutions developed in whichever program(s) applicants are comfortable using. Please include all the code snippets, and visualizations/screenshots in the PDF file. Do NOT attach any additional documents.

Questions?

If you require any clarifications on these instructions, please email dclarke@ea.com

Good Luck!

Take-Home Test

Context
At EA, the CDS Quality Verification team is an integral part of the game development process. From reviewing designs to verifying functionality and creating tools, the QV team works with their game development partners to create amazing games and services for our players.

One important function of QV is detecting software defects and then analyzing and documenting the defects, so they can be fixed quickly and precisely. We call the documentation for a given defect a "bug". All bugs are logged in bug tracking software where they are addressed through a standard workflow.

A typical bug could be associated with dozens of separate pieces of information - each providing data on different aspects of the bug including how severe the bug is, where in the game it occurs, how to reproduce the bug, and attachments such as screenshots, videos or other technical data. A QV data analyst's work is to analyze bug data.

With this test we have provided you with two csv files named 'Bugs' and 'BugHistory'. The 'Bugs' file is the latest record of bugs found in a particular project. Each given bug has a unique identifier called BugId with the latest information on Priority, Severity, Platform Found, Game Area, Test type, Progress Status, Created Time, and Closed Time. The 'BugHistory' file, on the other hand, contains historical information of each bug's progress status. Each record in 'Bughistory' provides the status of each bug captured at midnight on a given day.

Data Dictionary
The definitions of each field included in the two csv files have been sent to you via email.

Questions
1) Create a SQL query to merge the data from 'BugHistory' and 'Bugs' so that you get the count of severity-A Bugs for each 'FullDate'. Include only days with count of more than 7 bugs in the results.

2) The *Open Bugs KPI* represents the number of bugs with an open progress status. The statuses considered open are 'Claim Fixed', 'Confirmed', 'Fix Failed', 'In Progress', 'Info Added', 'Pending Build', 'Request Review', and 'Submitted'. This KPI helps us track the quality of our game and the productivity of the Development (Dev) and QV teams during the game development process. Using the data provided:

   a. Write a SQL query that shows the count of open bugs for each 'GameArea' and 'FullDate'.

    b. Using a program of your choice, provide a visualization of the Open Bugs KPI that shows the number of open bugs for each 'FullDate' broken down by 'GameArea'.

    c. Provide any insights that you can derive from this visualization on our game development process (150 words or less)

3) As described in the Data Dictionary, 'BugWeight' is a numerical value assigned to each bug for their quick and systemic prioritization. This value helps the Development (Dev) team decide which bugs should be resolved more quickly than others. The higher the Bug Weight, the more crucial it is for the Dev team to resolve the bug.

A Senior Data Analyst (DA) on the CDS team is interested in investigating the relationship of Bug Weight with Severity, Priority, User Path, and Repro Rate. The DA trains the following predictive model on a training set whose R output is given below:

Model:
BugWeight = a*Severity + b* UserPath + c* Priority + d*ReproRate + e
where a,b,c,d,and e are constants

R Output:

```
Call:
lm(formula = BugWeight ~ Severity + UserPath + Priority + ReproRate,
    data = training_set)

Residuals:
    Min      1Q  Median      3Q     Max
-316.57  -67.04   11.28   86.80  835.62

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   678.22     237.71   2.853 0.005708 **
Severity2    -391.53      84.15  -4.653 1.53e-05 ***
Severity3    -983.25      92.52 -10.628 3.57e-16 ***
UserPath2    -264.18      55.97  -4.720 1.19e-05 ***
UserPath3     439.76      64.54   6.814 2.89e-09 ***
Priority2     -26.63     209.42  -0.127 0.899190
Priority3     -46.28     200.88  -0.230 0.818461
Priority4     -90.50     200.76  -0.451 0.653536
Priority5      47.46     197.67   0.240 0.810959
ReproRate2    -65.70     128.70  -0.510 0.611357
ReproRate3    242.85     132.29   1.836 0.070710 .
ReproRate4    473.03     117.33   4.032 0.000141 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 191.7 on 69 degrees of freedom
Multiple R-squared:  0.8421,         Adjusted R-squared:  0.8169
F-statistic: 33.46 on 11 and 69 DF,  p-value: < 2.2e-16
```

Using the R output, answer the following questions to help the DA in their investigation:

a. What measure from the above output would you recommend using to compare this model to another model of the same type with different variables?
b. What are the underlying assumptions that must be made for this model to be true?
c. From the above R output, which variable, if any, has the strongest impact on 'BugWeight' and why? (50 words or less)