Code.org Data Analyst Take-Home Assignment

Overview

This exercise is a way to see how you think and to gauge your comfort with the types of skills and tools you will use in the role.

In total, this exercise should take you no more than 2-3 hours to complete. *We don't expect an exhaustive analysis in that limited time - we're primarily focused on the question you ask and how you approach the data.*

How to submit your responses

Please upload any material for your responses using the link provided in the instructions email.

- **IMPORTANT: DO NOT include your name anywhere in the submission.**
  The link to which you are uploading your materials is unique to you. We review all submissions anonymously to help ensure fairness in our hiring process. Thank you for helping us with this process.

Exercise Details

Scenario

A Code.org leader who is focused on increasing the high school adoption of our curriculum has asked you for insights on where to target their efforts in the coming year. How would you use the data available to answer this question?

1) **First, define your central research question.** There are many ways to approach this question - how will you do so with the data you have? Remember to take into account the organization's mission.
    a) Example question: "How has the usage of Code.org's high school curriculum changed over time?"
2) **Generate 2-3 user-facing charts or visualizations that explore the data to answer that question**
3) **Synthesize the data and write up your conclusions**
    a) What key takeaways would you give to leadership? (bullet points are fine)
    b) If you had more data and a longer timeframe, what other data would you want to look at? What other questions would you ask?

This exercise should take you 2-3 hours in total.

Reminder: There are no right or wrong answers to this question and we do not expect an exhaustive analysis. We want to understand how you approach an analysis task.

The Data

You should have access to a sheet called High School Adoption 2021-24

The data shows which high schools are using one or more of Code.org's high school curricula over the past 3 years. The relevant curricula are: 1) AP CS Principles (csp); 2) AP Computer Science A (csa); and 3) CS Discoveries (csd), a middle school-aligned course that is often used at high schools; and 4) a non-AP "special topics" course on computer vision + AI (9-12 special topics). The data are represented as one row per school_id/school_year. Each school will have three rows in the data, one row per school year.

Here is a reference for the fields included in the dataset.

Field	Description

school_id	Unique school identifier Note that these IDs are based on NCES IDs but have been masked to allow us to share data for this exercise.
school_year	School year starting July 1st of a calendar year and ending June 30th of the following calendar year
active_courses	Whether the school meets our internal bar for implementing a course in the classroom. This is based on a number of metrics, one of which is a requirement of at minimum 5 participating students. This field contains a comma separated list of values. Note that many of these schools teach other Code.org curriculum, but this list has been limited to only the curricula listed above: csa, csp, csd, and 9-12 special topics.
state	US State where the school is located
is_stage_el is_stage_mi is_stage_hi	Whether the school offers grades K-5, 6-8, and/or 9-12 (1- Yes, 0 - No) This dataset is limited to schools that offer (at least) 9-12
is_rural	Indicator for schools located in rural communities (1 - Yes, 0 - No)
is_title_i	Whether the school is considered Title I (1 - Yes, 0 - No)
community_type	The type of community in which the school is located
school_type	Type of school (public, private, or charter)
is_high_needs	Indicator for schools where at least 50% of the students are eligible to receive free or reduced lunch (1 - Yes, 0 - No)
total_students	The total # of students enrolled at the school
urg_percent	The percentage of enrolled students at the school who come from racial and ethnic backgrounds traditionally underrepresented in computer science (1 - Yes, 0 - No)
frl_eligible_percent	The percentage of students that are eligible for free or reduced lunch
num_pd_teachers	The number of teachers who have received substantial PD from Code.org *that school year* - note that this includes either attending at least one in-person workshop or significant use of our self-paced curriculum for teachers
num_csp_students num_csa_students num_csd_students num_special_topics_students	The number of students at the school who engaged with our AP CSP, AP CSA, CS Discoveries, or 9-12 special topics content that school year

Submission

Please upload any outputs you create to the submission link included in the instructions email.

If you write code, please include that too. Note that the link allows most file types but does **not** allow executable files such as .exe, .py, or .r. If you have trouble uploading an executable file to the link provided, please email it directly to tj@code.org.

How we'll evaluate it:

- We want to see a clearly defined research question at the center of your analysis.
- Keep your analysis simple, clear, and direct. (If you end up with a lot of files, a ReadMe is appreciated!)
- Use whatever tools you prefer to do the analysis work. There is no need for a formal presentation or slides - a spreadsheet is fine. If you do choose to write code, please include it.
- We're looking for clarity of thought, comprehensiveness, and a strong analytical approach.

Deadline and next steps

You should have received an email with a specific date and time deadline but you may submit responses before the deadline. If you need any help with this or have any questions please feel free to reach out.

All submissions are **anonymized** before the hiring team does their review. RED team members will review your work independently and score it against a rubric, then meet to discuss.

From this work, we may decide to move forward with a full interview loop, pause, or discontinue your candidacy. You should hear back from us within 10 days of the homework deadline.