The importance of validating a dataset

You work for a company that helps maintain the data of an international organization that manages and hosts long-distance running competitions. Your manager has just assigned you to validate the data input gathered from the entrants for a recent marathon held in Cyprus. Runners from all over the world traveled to compete in the race. Your manager tells you that the data has already been cleaned, and it is now ready for final input validation.

When you take a look at the data in Python, you find the following columns titles:

Entrant Name	Sign-Up Date	Age Group	Completion Time	Home Country	Entry Fee Amount Paid

To complete this discussion prompt:

. Select two of the data columns above (other than “entrant name,” which has examples provided below).
. Write what you would validate in the two data columns you selected. For example, for “entrant name”:

Format: first name, last name, plus initials

Script: names in English or other languages/scripts

Duplicates: check that no one entered twice or more

Missing/fake names: search for only first or last names, or fake entries

. List three to four things to validate for each column (60-80 words).