Congratulations! You passed!
Grade received 100%To pass 100% or higher

1. Activity overview

So far, you've learned a lot about the importance of cleaning data and how to do it in spreadsheets and SQL. In this activity, you'll follow a scenario and clean real data in R.
By the time you complete this activity, you will learn more about data cleaning functions in R and apply this know-how to import, preview, and perform calculations on different data sets. You can use these techniques to gain initial insights into your data, which will help you analyze data throughout your career.

Working in RStudio Cloud

To start, log into your RStudio (Posit) Cloud account. Open the project you will work on in the activity with this link, which opens in a new tab. If you haven't gone through this process already, at the top right portion of the screen you will see a "red stamp" indicating this project as a Temporary Copy. Click on the adjacent button, Save a Permanent Copy, and the project will be saved in your main dashboard for use with future lessons. Once that is completed, navigate to the file explorer in the bottom right and click on the following: Course 7 -> Week 3 -> Lesson3_Clean.Rmd.
The .csv file, hotel_bookings.csv, is also located in this folder.

If you have trouble finding the correct activity, check out this step-by-step guide on how to navigate in RStudio (Posit) Cloud. Make sure to select the correct R markdown (Rmd) file. The other Rmd files will be used in different activities.
If you are using RStudio Desktop, you can download the Rmd file and the data for this activity directly here:

Lesson3_Clean
RMD File

hotel_bookings
CSV File

You can also find the Rmd file with the solutions for this activity here:

Lesson3_Clean_Solutions
RMD File

Carefully read the instructions in the comments of the Rmd file and complete each step. Some steps may be as simple as running pre-written code, while others may require you to write your own functions. After you finish the steps in the Rmd file, return here to confirm that your work is complete.

Confirmation

In Step 5 of this activity, you created the number_canceled column to represent the total number of canceled bookings. What value was returned in this column?

- 40234
- 44224
- 49550
- 52965

Correct
The number returned in this number_canceled column should be 44,224, which represents the total number of canceled hotel bookings. By cleaning and manipulating the data, you were able to answer an important question about it. Going forward, you can use what you know about data cleaning from past courses to help you learn how to clean data in R.