User Study

Case1:

Question:

A developer wants to filter the data. He only wants to keep the data where the airline is airline_1. He also wants to drop duplicate data. Finally, he counts the number of data. There may be a problem with his code, please run the code. If there is a problem, please try to describe the problem and reason.

Case2:

Question:

There are a total of 18 columns (named column1, column2, …, column18). A developer wants to delete 8 columns. Please run the code, and if there is a problem, please try to describe the problem and reason.

Case3:

Question:

A developer wants to get the data of the passengers older than 80 years old. As there may be duplicate passenger information, he wants to remove the duplicate data. Finally, he counts the number of selected and remaining data. The following is his code. Note that the data is imported automatically from pandas. Please run the code, and if there is a problem, please try to describe the problem and reason.

Case 4:

Question:

We want to replace all null values (NA) with an integer 10. After that, there should be no null value in the data. Please run the code, and if there is a problem, please describe the problem and reason.

Case 5:

Question:

A developer wants to create a new column based on the original column. The values of the original ticket price should be all in the range from 0$ to 100 $. The value of the new ticket price is increased by 50$ based on the original value. The range of the new column value should be 50 $-150 $. The developer will use the new column mean value as the general airfare price for advertising. Please run the code, and if there is a problem, please describe the problem and reason.

Case 6: A developer wants to delete rows with null values in the data. Then he wants to know how many passengers are female. Please run the code, and if there is a problem, please describe the problem and reason.