How to transform the training data into test data format:

Two suggested approaches

A. First, we are trying to add a column to the raw data, that indicates for each row, for which time interval (which is represented by a single timepoint at the beginning of the time interval) it belongs. We can do it with the following steps:

1. Create a vector containing all relevant timepoints (i.e. every 15 minutes throughout the time period you have selected).

2. Use int_diff function from lubridate package on this vector to create an interval object.

3. Run a loop on all the time points in the raw data and for each data point; indicate to which interval (which is represented by a single timepoint at the beginning of the time interval) it belongs. Put the results of it in a new column in the raw data.
   Hint: %within% argument might help.

Now use dplyr to group by the new column you have added, and find the appropriate summarise function to count how many rides were in each time interval.

B. This approach contains the following steps:

1. Create a vector containing all relevant timepoints (i.e. every 15 minutes throughout the time period you have selected) and name it time_intervals (for example)

2. Run a for loop on time_intervals, and in each iteration, run another for loop on all the time points in the raw data. For each time point, indicate whether it less than 15 minutes after the specific value of time_intervals. Count how many rides were in each time interval.
   Hint: the function ‘difftime’ may help you calculate time differences.

3. Create a data frame containing the result of the double loop and time_intervals.