Data Wrangling

- First, both the train & test data was imported and merged, so that all the operations can be performed on the entire dataset. In total it had 275547 users' activity data.
- Once the structure of the dataset was clear all the missing values were checked. 46% users did not disclose their gender.
- Similarly, there were 42% of users without any age information. In addition to that, I came across a few outliers in the age column. 830 of the users had mentioned their age of more than 115 years. Which I excluded and kept the data with the age range of 13 to 95 years.
- I check the 'country_destination' column against the gender column. But could not find any noticeable difference between both the genders.
- The first thing I could see that if there was a reservation, it was likely to be inside the US. But there were around 45% of people that never did any reservation.
- After that, I compared it against age and came to know that the common age to travel was between 25 and 45.