# User Behavior Data for Datathon

You are provided browsing behavior of website visitors who are shopping for a hotel. Typically, a user visits a travel website several times for research before they make a final booking. It is important to understand at what stage of the booking funnel the user is in so that we can personalize and optimize the website to the user's interest. For example, is the user still in research mode trying to choose the right destination (early in booking funnel) vs. the user made decision on a specific hotel and browsing for the best price (bottom of the booking funnel). The type of information a user is looking for is completely different for these two use cases and it is very impactful if we personalize the website accordingly.

In the given data set, we have aggregated the past browsing history of the user into several different features. The objective of the datathon is to predict the probability of a hotel booking in their next visit (how close the user is to making a final booking purchase). Note that for confidentiality reasons, the column names are masked or renamed without any loss of interesting insights for data science. Also the data is a skewed sample of millions of users and may not represent the typical traffic to a travel website. Here are the details of the data set. All the columns with the name starting with '*p_*' are based on the user previous browsing history. If any of the data is 'NA', assume the data is not available or unknown.

| Column | Description |
|---|---|
| *user_id* | Unique user id |
| *day* | Day of the user visit to the website |
| *gender* | User gender |
| *p_sessionActivity* | A metric that measures how active the user is on website during the past visit |
| *p_AddToCart* | User added the hotel to the cart during the past visit |
| *p_trafficChannel* | Marketing channel of the user's past visit |
| *p_sessionDuration* | Time spent by user on the website during the past visits (seconds). |
| *p_pageViews* | Numbers of pages browsed by the user during the past visits |
| *daysToCheckin* | Number of days to checkin date. Assume the data is not available when it is NA |
| *osType* | Operation system type |
| *osTypeName* | Operating system name |
| *daysFromPreviousVisit* | Number of days from the latest past visit |
| *p_TotalPrice* | Total max price of hotels the user looked at in the past visit. NA implies not available. |
| *isExclusiveMember* | User is an exclusive member |
| *loggedIn* | User has an account and logged in |
| *p_MapInteraction* | Interaction with maps on the website |
| ***BookingPurchase*** | **User made a purchase. We need to predict this based on the above data** |