Team Offset - DataFest

We attempted to correlate weather data with booking rates by using external weather api to get weather metadata for the checkin date and the lat/lng coordinates of the destination. We mapped the number of bookings vs the each type of weather data and found that clear-day had the highest count of booking which made sense until we looked at the average rate of bookings per weather type and we found the distribution to be found all over the place. This indicated that the weather did not make much impact on the booking rates. For the dates that there were bookings and the dates that did not having bookings, we looked at the average min/max temperature. We found that the min temperature and the max temperature for days that people booked and the days that people didn't book were almost equal. This confirmed our suspicion that the weather did not make a big impact on the booking rates. We also tried to use machine learning to figure out which features were important in predicting if a user will book or not. We found that the features **channel** and **cnt** were important in predicting if a user will book or not.

While we were doing our analysis with weather, we realized that some of the checkout dates were before checkin dates. We ran this analysis on the big data set and found that 1755 rows had this issue. Not only that, some of the rows had no checkin dates or checkout dates or some of these dates were in the wrong format. This is just something interesting that we found in the dataset.

We also did some hit rates for booking. We analyzed the influence of 3 key predictors (international, prop_is_branded, cnt) to determine their impact on the probability that a use books a hotel room(is_booking). We found that travelling domestically always has higher hit rates for booking than travelling internationally. Also branded always has a higher hit rates for booking than unbranded, and the branded hit rate is nearly equal to the domestic hit rate. Another trend we found which is somewhat intuitive is that multiple clicks lead to bookings much less often than single clicks.